

Programación Python para BigData

Lección 4: Base de datos: PostgreSQL





ÍNDICE

Le	Lección 4: Base de datos: PostgreSQL2				
Presentación y objetivos					
	PostgreSQL				
2.	Adminer	13			
3.	Psycopg2	23			
4.	Puntos claves	32			



Lección 4: Base de datos: PostgreSQL

PRESENTACIÓN Y OBJETIVOS

En esta lección aprenderemos a trabajar con un ejemplo de base de datos relacionable como es PostgreSQL, usando tres instrumentos para saber cómo leer, insertar, actualizar o eliminar un dato. (CRUD)



Objetivos

- Trabajar a través de consola usando directamente PostgreSQL
- Trabajar con Adminer es un visualizador de la base de datos.
- Trabajar con una librería de Python como es Psycopg2



1. PostgreSQL

Introducción base de datos relacionales

La base de datos relacional (BDR) es un tipo de base de datos (BBDD) que cumple con el modelo relacional (el modelo más utilizado actualmente para implementar las BBDD ya planificadas). Todos los datos se almacenan y se accede a ellos por medio de relaciones previamente establecidas.

Las relaciones que almacenan datos son llamadas relaciones base y su implementación es llamada "tabla".

Otras relaciones no almacenan datos, pero son calculadas al aplicar operaciones relacionales. Estas relaciones son llamadas relaciones derivadas y su implementación es llamada "vista" o "consulta". Las relaciones derivadas son convenientes, ya que expresan información de varias relaciones actuando como si fuera una sola tabla.

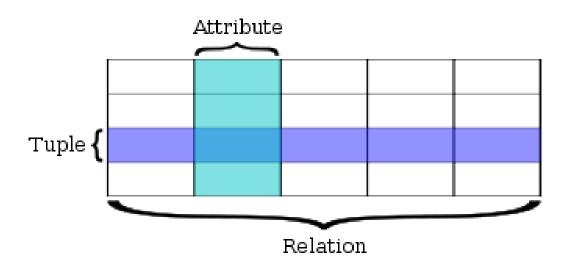


Figura 1.1 Estructura de la tabla/relación en SQL. Fuente Wikipedia



Término SQL	Término de bases de datos relacionales	Descripción
Fila	Tupla o registro	Un conjunto de datos, que representa un ítem simple
Columna	Atributo o campo	Un elemento etiquetado de una tupla, p.e. "Nombre" o "Fecha"
Tabla	Relación	Un conjunto de tuplas compartiendo los mismos atributos; un conjunto de filas y columnas.

Clasificación de Claves

- Clave primaria o Private key: Una clave primaria es una clave única (puede estar conformada por uno o más campos de la tabla) elegida entre todas las candidatas que define unívocamente a todos los demás atributos de la tabla para especificar los datos que serán relacionados con las demás tablas. La forma de hacer esto (relación entre tablas) es por medio de claves foráneas.
- Clave externa o foránea o Foreign key: Una clave foránea es una referencia a una clave en otra tabla, determina la relación existente en dos tablas. Las claves foráneas no necesitan ser claves únicas en la tabla donde están y sí a donde están referenciadas.
- Clave índice o ID: Las claves índice surgen con la necesidad de tener un acceso más rápido a los datos. Los índices pueden ser creados con cualquier combinación de campos de una tabla. Las consultas que filtran registros por medio de estos campos, pueden encontrar los registros de forma no secuencial usando la clave índice.

Las bases de datos relacionales incluyen múltiples técnicas de ordenamiento, cada una de ellas es óptima para cierta distribución de datos y tamaño de la relación.



Los índices generalmente no se consideran parte de la base de datos, pues son un detalle agregado. Sin embargo, las claves índices son desarrolladas por el mismo grupo de programadores que las otras partes de la base de datos.

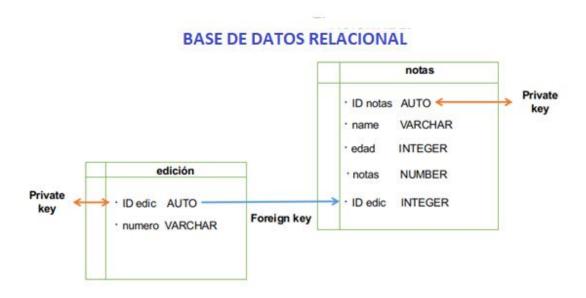


Figura 1.2 Estructura de relación entre tablas en SQL.

Introducción a PostgreSQL

PostgreSQL, o conocido como Postgres es un sistema de código abierto de administración de bases de datos del **tipo relacional o SQL**, aunque también es posible ejecutar consultas que sean no relaciones. En este sistema, las consultas relacionales se basan en SQL, mientras que las no relacionales hacen uso de JSON.

Es gratuito, y su desarrollo es llevado adelante por una gran comunidad de colaboradores de todo el mundo para hacer de este sistema una de las opciones más sólidas a nivel de bases de datos.

Dos detalles a destacar de PostgreSQL es que posee data types (tipos de datos) avanzados y permite ejecutar optimizaciones de rendimiento avanzadas, que son características que por lo general solo se ven en sistemas de bases de datos comerciales, como por ejemplo SQL Server de Microsoft u Oracle de la compañía homónima.



Características

- Es de código abierto.
- Es gratuito.
- Es multiplataforma: es un software que puede correr bajo distintos entornos y sistemas operativos.
- Es fácil de usar
- Puede manejar un gran volumen de datos.
- Soporte total de ACID: siglas de Atomicity, Consistency, Isolation y Durability, que si lo traducimos al español básicamente hablan de la atomicidad, consistencia, aislamiento y durabilidad de las transacciones que se realizan en una base de datos. Esto es importante ya que da la seguridad de que, si se produce una falla durante una transacción, los datos no se perderán ni terminarán donde no deban.

Comparación de PostgreSQL vs MySQL

PostgreSQL es un sistema de código libre y totalmente gratuito que está en manos de una comunidad. Con MySQL sigue existiendo una versión gratuita aunque eso solo será hasta que Oracle así lo quiera, ya que hace unos años pertenece a ellos. Afortunadamente existen alternativas MySQL extremadamente similares como MariaDB, por si algún día necesitas un reemplazo.

En lo que refiere a las consultas o queries, las soportadas por MySQL tienden a ser más simples, mientras que PostgreSQL soporta queries más complejas, lo cual le ha valido el título del sistema de bases de datos relacionales más avanzado del mercado.



Rendimientos:

MySQL es mejor si se manejan bajos volúmenes de datos, es decir, si tenemos bases de datos pequeñas o medianas a las cuales se hagan una cantidad de consultas no muy alta.

PostgreSQL es mejor a la hora de manejar volúmenes de datos grandes y se suele usar más cuando se tienen bases de datos grandes y con alta cantidad de consultas.

Cumplimiento de ACID:

MySQL no lo soporta al 100%, sino solamente las tablas que hagan uso de NDB Cluster o InnoDB.

PostgreSQL sí lo soporta completo en todos los casos.

Por todo ello como nuestra asignatura es Big Data usaremos como base de datos PostgreSQL

Ejecutar PostgreSQL

Realizaremos las actividades descritas como **CRUD**, que viene del Inglés (Create, Read, Update and Delete):

Create / Crear: crearemos un nuevo dato en nuestra tabla.

Read / leer: leeremos los datos de la tabla.

Update / actualizar: actualizaremos un dato ya existente en la tabla.

Delete / eliminar: eliminamos un dato de la tabla.

Para ello usamos el docker-compose creado en la actividad de la lección 2, donde tendremos activos los servicio de PostgreSQL y adminer para la visualización y gestión de los datos.

Necesitamos para trabajar con comandos en el PostgreSQL ejecutar la imagen:



```
tsabel@isabel-SVE1512E1EM:~/atom$ sudo docker ps
CONTAINER ID IMAGE COMMAND CREATED STATUS PORTS NAMES
473837faeb01 adminer "entrypoint.sh docke..." 8 minutes ago Up 36 seconds 0.0.0.0:8080->8080/tcp, :::8080->8080/tcp atom_adminer_1
9ce68f58dbe4 postgres "docker-entrypoint.s..." 8 minutes ago Up 36 seconds 0.0.0.0:5432->5432/tcp, :::5432->5432/tcp atom_db_1
```

Figura 1.3 Imágenes ejecutadas a través de docker-compose.

```
sudo docker exec -it ato_db_1 bash
psql -U IsaMan
```

-U es el usuario descrito en la imagen al ejecutar el docker-compose

Figura 1.4 Ejecución de PostgreSQL

Creamos una base de datos

```
CREATE DATABASE <nombre de la base de datos>
```

```
IsaMan=# CREATE DATABASE test;
CREATE DA<mark>T</mark>ABASE
```

Figura 1.5 Crear un base de datos en PostgreSQL

Eliminar una base de datos

```
DROP DATABASE <nombre de la base de datos>
```



IsaMan=# DROP DATABASE test; DROP DATABASE IsaMan=#

Figura 1.6 Eliminar base de datos en PostgreSQL

Crear una tabla en la base de datos

```
CREATE TABLE <nombre_tabla>(<nombre_columna> <tipo_variable>,...);
```

```
IsaMan=# CREATE TABLE notas(name varchar(80), age int, grades decimal, date date);
CREATE TABLE
```

Figura 1.7 Crear una tabla en la base de datos en PostgreSQL

		Tipos de Variables
	bigint (tamaño)	Número de gran tamaño entero
	int (tamaño)	Número entero
	smallint(tamaño)	Número entero pequeño tamaño
Número	real	También decimal. Datos numéricos de precisión flotante
	money	Datos monetarios
	boolean	Booleano significa si/no, on/off, etc
	numeric	Numérico
	double precision	Decimal con precisión
	date	Una fecha. Formato: AAAA-MM-DD
Fecha y hora	time	Un tiempo. Formato: HH: MI: SS
reciia y ilora	timestamp	Marca de tiempo en segundos
	timestampz	Marca de tiempo en segundos UTC
	interval	Intervalo de tiempo



	Character(tamaño)	o también llamado char.cadena de caracteres
	Character varying(tamaño)	o también llamado varchar. cadena de caracteres de ancho variable
	text	Cadena de caracteres de ancho variable
	ts query	Almacena en formato cursor
Cadena	ts vector	Almacena en formato vector
	uuid	Almacenar en formato UUID
	json	Almacenar en formato JSON
	xml	Almacenar en formato XML
	jsonb	Almacenar en formato BSON
Divavia	bit	Entero que puede ser 0, 1 o NULL
Binario	bit varying	Entero que puede ser variable
	bytea	cadena de caracteres binaria
	cidr	
Red	inet	
	marcaddr	
	txid_snapshot	
	box	Formato caja
	circle	Formato de circulo
Geometría	line	Formato de línea
Geometria	lseg	
	path	Formato de ruta
	point	Formato de puntos
	polygon	Formato de Polígono

Insertar dato

INSERT INTO <nombre tabla> VALUES (<valores>)



INSERT INTO notas VALUES ('Isabel Maniega', 205, 5.4, '07/07/2021')

```
IsaMan=# INSERT INTO notas VALUES ('Isabel Maniega', 205, 5.4, '07/07/2021');
INSERT 0 1
IsaMan=#
```

Figura 1.8 Insertar un datos en la tabla en la base de datos en PostgreSQL



NOTA: La fecha en formato americano (MES/ DIA/ AÑO)

Leer datos

```
SELECT * FROM <nombre_tabla>;
```

* Significa seleccionar todas las columnas de la tabla.

También se puede poner en vez de * el nombre de la columna/s SELECT * FROM notas;

Figura 1.9 Seleccionar datos en la tabla en la base de datos en PostgreSQL

Actualizar datos

```
UPDATE <nombre_tablas> SET columna=valor_a_modificar;
```

UPDATE notas SET grades=6.9;

```
IsaMan=# UPDATE notas SET grades=6.9;
UPDATE 1
IsaMan=#
```

Figura 1.10 Actualizar datos en la tabla en la base de datos en PostgreSQL



Eliminar datos

DELETE FROM <nombre_tabla> WHERE columna=valor;

DELETE FROM notas WHERE name='Isabel Maniega'

```
IsaMan=# DELETE FROM notas WHERE name='Isabel Maniega';
DELETE 1
IsaMan=#
```

Figura 1.11 Eliminar datos en la tabla en la base de datos en PostgreSQL

Eliminar tabla

DROP TABLE <nombre_tabla>;

IsaMan=# DROP TABLE notas; DROP TABLE

Figura 1.12 Eliminar tabla en la base de datos en PostgreSQL



2. ADMINER

Adminer es una aplicación para gestionar base de datos SQL, como hemos usado docker-compose según la descripción de la imagen PostgreSQL, viene el servicio Adminer que usaremos para gestionar la base de datos usando una aplicación.

Para ello sólo necesitamos ir al navegador y poner:

http://localhost:8080/

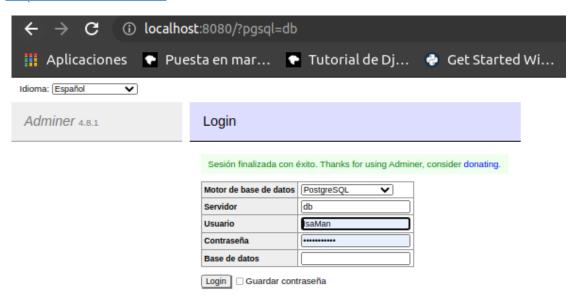


Figura 2.1 Pantalla inicial de la aplicación Adminer

Motor de base de datos: PostgreSQL

Servidor: db

Usuario: el puesto en docker-compose

Contraseña: el puesto en docker compose

Base de datos: vacio

Pulsamos Login





Figura 2.2 Pantalla donde podemos ver nuestras bases de datos activas en PostgreSQL

Creamos una base de datos

Seleccionamos crear Base de datos:



Figura 2.3 Crear bases de datos en PostgreSQL





Figura 2.4 Nombrar la base de datos en PostgreSQL

Ponemos "eip" y pulsamos Guardar.

Crear una tabla en la base de datos



Figura 2.5 Base de datos "eip" en PostgreSQL

Pulsar crear tabla:



Figura 2.6 Tabla notas en base de datos "eip" en PostgreSQL



Pondremos:

Nombre de la tabla: notas

Name: character 80

Age: integer

Grades: double precision

Date: date

Pulsaremos Guardar:



Figura 2.7 Tabla notas en base de datos "eip" en PostgreSQL

Ya nos ha creado la nueva tabla.



Insertar datos

Pulsamos nuevo registro:



Figura 2.8 Insertar datos en tabla notas en base de datos "eip" en PostgreSQL



Figura 2.9 Insertar datos en tabla notas en base de datos "eip" en PostgreSQL

Insertamos los datos como la imagen y pulsamos guardar:





Figura 2.10 Insertar datos en tabla notas en base de datos "eip" en PostgreSQL

Podemos ver los datos que hemos introducido.

Actualizar un dato



Figura 2.11 Modificar datos en tabla notas en base de datos "eip" en PostgreSQL

Pulsaremos modify:





Figura 2.12 Modificar datos en tabla notas en base de datos "eip" en PostgreSQL

Modificamos la nota y pulsamos

Probamos con comandos:



Figura 2.13 Modificar datos en tabla notas en base de datos "eip" en PostgreSQL

UPDATE notas SET grades=7.9 WHERE name='Isabel Maniega';



Ejecutamos:



Figura 2.14 Modificar datos en tabla notas en base de datos "eip" en PostgreSQL

Pulsamos Modificar:

Si volvemos a la página de notas a Visualizar contenido vemos que nos lo ha modificado:



Figura 2.15 Modificar datos en tabla notas en base de datos "eip" en PostgreSQL



Eliminar un dato

Seleccionamos el dato y pulsamos eliminar nos aparece un aviso si estamos seguros, damos aceptar:



Figura 2.16 Eliminar datos en tabla notas en base de datos "eip" en PostgreSQL

Por comandos sería:



Figura 2.17 Eliminar datos en tabla notas en base de datos "eip" en PostgreSQL



DELETE FROM notas WHERE name='Isabel Maniega';

Pulsamos ejecutar



Figura 2.18 Eliminar datos en tabla notas en base de datos "eip" en PostgreSQL

Pulsaremos después a Modificar.

Volvemos a la ventana principal a Visualizar contenido:



Figura 2.19 Eliminar datos en tabla notas en base de datos "eip" en PostgreSQL



3. Psycopg2

Librería de Python para gestión de base de datos en PostgreSQL.

Para ello necesitamos crear el entorno virtual:

```
isabel@isabel-SVE1512EIEW:-/atom/postgres$ virtualenv env
created virtual environment CPython3.8.10.final.0-64 in 294ms
creator (CPython3Postx/dest=/home/isabel/atom/postgres/env, clear=False, global=False)
seeder FromAppData(download=False, pip=latest, setuptools=latest, wheel=latest, pkg_resources=latest, via=copy, app_data_dir=/home/isabel/.local/sha
re/virtualenv/seed-app-data/v1.0.1.debian.1)
activators Bashactivator, CShellActivator, FishActivator, PowerShellActivator, PythonActivator, XonshActivator
isabel@isabel-SVE1512EIEW:-/atom/postgres$.
```

Figura 3.1 Creación de entorno virtual

En atom creamos el archivo "postgres.py":

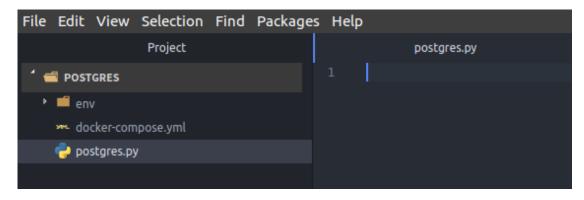


Figura 3.2 Creación de archivo Python

Activamos el entorno e instalamos la librería:

```
pip install psycopg2-binary
```

Figura 3.3 Creación de archivo Python

Para conectarnos a la base de datos ponemos:



```
postgres.py

import psycopg2

# Connect to an existing database
conn = psycopg2.connect(user='IsaMan', password='cursoPython',
host='localhost', port=5432)

# Open a cursor to perform database operations
cur = conn.cursor()
```

Figura 3.4 Conexión a la base de datos

Crear una base de datos nueva

Añadimos las siguientes librerías:

```
from psycopg2 import sql
from psycopg2.extensions import ISOLATION_LEVEL_AUTOCOMMIT
```

Colocamos entre la conexión y el cursor:

```
conn.set_isolation_level(ISOLATION_LEVEL_AUTOCOMMIT)
```

Después creamos la base de datos:



Figura 3.5 Creación de base de datos llamada "eip"

```
(env) isabel@isabel-SVE1512E1EW:~/atom/postgres$ python postgres.py
(env) isabel@isabel-SVE1512E1EW:~/atom/postgres$
```

Figura 3.6 Ejecutar el script



Figura 3.7 Visualización de la base de datos "eip" en Adminer

Observamos que se ha creado.



Crear una tabla

Una vez creada creamos una tabla nueva, debemos antes de seguir comentar la parte de crear la base de datos, para no tener problemas con el código.

Añadimos en conn la variable database donde vamos a conectar y la parte para crear la tabla:

Añadimos al final:

```
conn.commit()
```

Para hacer que los cambios sean persistentes

```
import psycopg2

# Connect to an existing database
conn = psycopg2.connect(database="eip", user='IsaMan', password='cursoPython',
host='localhost', port=5432)

# Open a cursor to perform database operations
cur = conn.cursor()

# Creamos una nueva tabla:
try:
cur.execute("CREATE TABLE notas (name varchar(80), age int, grades real, date date);")
except psycopg2.Error as e:
print('Error Crear tabla: %s', str(e))

# Close communication with the database
conn.commit()
cur.close()
conn.close()
```

Figura 3.8 Creación de la tabla en base de datos "eip"





Figura 3.9 Visualización de la tabla notas en la base de datos "eip" en Adminer

Insertar un dato

Una vez creada la tabla, debemos antes de seguir comentar la parte de crear la tabla, para no tener problemas con el código.

Pondremos:



Figura 3.10 Insertar un dato en la tabla notas en la base de datos "eip"



Figura 3.11 Visualización de la tabla notas en la base de datos "eip" en Adminer



Lectura de los datos

```
# Query the database and obtain data as Python objects
cur.execute("SELECT * FROM notas;")
rows = cur.fetchall()
for row in rows:
    print(row)
```

Figura 3.12 Lectura datos en la tabla notas en la base de datos "eip"

Actualizar un dato

```
# actualizar datos buscando el dato a actualizar
try:
    cur.execute("UPDATE notas SET grades=7.9 WHERE
name='Isabel Maniega'")
except psycopg2.Error as e:
    print('Error Actualizar dato: %s', str(e))
```



```
import psycopg2

# Connect to an existing database

conn = psycopg2.connect(database="eip", user='IsaMan', password='cursoPython',

host='localhost', port=5432)

# Open a cursor to perform database operations

cur = conn.cursor()

# actualizar datos buscando el dato a actualizar

try:

cur.execute("UPDATE notas SET grades=7.9 WHERE name='Isabel Maniega'")

except psycopg2.Error as e:

print('Error Insertar dato: %s', str(e))

# Query the database and obtain data as Python objects

cur.execute("SELECT * FROM notas;")

rows = cur.fetchall()

for row in rows:

print(row)

# # Make the changes to the database persistent

conn.commit()

# Close communication with the database

cur.close()

conn.close()
```

Figura 3.13 Actualizar datos en la tabla notas en la base de datos "eip"

```
(env) isabel@isabel-SVE1512E1EW:~/atom/postgres$ python postgres.py ('Isabel Maniega', 205, 7.9, datetime.date(2021, 7, 7))
```

Figura 3.14 Salida consola: actualizar datos en la tabla notas en la base de datos "eip"

Eliminar los datos

```
try:
    cur.execute("DELETE FROM notas WHERE name='Isabel
Maniega';")
except psycopg2.Error as e:
print('Error eliminar dato: %s', str(e))
```



```
import psycopg2

# Connect to an existing database
conn = psycopg2.connect(database="eip", user='IsaMan', password='cursoPython',
host='localhost', port=5432)

# Open a cursor to perform database operations
cur = conn.cursor()

try:
cur.execute("DELETE FROM notas WHERE name='Isabel Maniega';")
except psycopg2.Error as e:
print('Error eliminar dato: %s', str(e))

# Query the database and obtain data as Python objects
cur.execute("SELECT * FROM notas;")
rows = cur.fetchall()
for row in rows:
print(row)
# # Make the changes to the database persistent
conn.commit()

# Close communication with the database
cur.close()
conn.close()
```

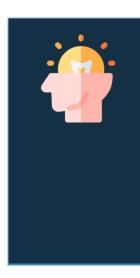
Figura 3.15 Eliminar datos en la tabla notas en la base de datos "eip"

```
(env) isabel@isabel-SVE1512E1EW:~/atom/postgres$ python postgres.py
('José Manuel Peña', 215, 8.6, datetime.date(2021, 7, 7))
```

Figura 3.16 Salida consola: Eliminar datos en la tabla notas en la base de datos "eip"



4. PUNTOS CLAVES



No te olvides...

- PostgreSQL es un base de datos de tipo relacionales o SQL.
- Para visualizar los datos de manera fácil usaremos Adminer.
- Psycopg2 es una librería de Python que nos facilita la gestión de la base de datos.

